Studying with the Help of Digital Tutors: Design Aspects of Conversational Agents that Influence the Learning Process

Wellnhammer, Natalie ; Dolata, Mateusz ; Steigler, Susanne ; Schwabe, Gerhard

Abstract: Conversational agents such as Apple's Siri or Amazon's Alexa are becoming more and more prevalent. Almost every smart device comes equipped with such an agent. While on the one hand they can make menial everyday tasks a lot easier for people, there are also more sophisticated use cases in which conversational agents can be helpful. One of these use cases is tutoring in higher education. Several systems to support both formal and informal learning have been developed. There have been many studies about single characteristics of pedagogical conversational agents and how these influence learning outcomes. But what is still missing, is an overview and guideline for atomic design decisions that need to be taken into account when creating such a system. Based on a review of articles on pedagogical conversational agents, this paper provides an extension of existing classifications of characteristics as to include more fine-grained design aspects.

Originally published at:
Wellnhammer, Natalie; Dolata, Mateusz; Steigler, Susanne; Schwabe, Gerhard (2020). Studying with the Help of Digital Tutors: Design Aspects of Conversational Agents that Influence the Learning Process. In: Proceedings of the 53rd Hawaii International Conference on System Sciences, Maui, Hawaii, USA, 7 January 2020 - 10 January 2020. University of Hawai'i at Manoa, 146-156.

Studying with the Help of Digital Tutors: Design Aspects of Conversational Agents that Influence the Learning Process

Natalie Wellnhammer
University of Zurich
natalie.wellnhammer@uzh.ch

Mateusz Dolata
University of Zurich
dolata@ifi.uzh.ch

Susanne Steigler
University of Zurich
steigler@ifi.uzh.ch

Gerhard Schwabe
University of Zurich
schwabe@ifi.uzh.ch

Abstract

Conversational agents such as Apple's Siri or Amazon's Alexa are becoming more and more prevalent. Almost every smart device comes equipped with such an agent. While on the one hand they can make menial everyday tasks a lot easier for people, there are also more sophisticated use cases in which conversational agents can be helpful. One of these use cases is tutoring in higher education. Several systems to support both formal and informal learning have been developed. There have been many studies about single characteristics of pedagogical conversational agents and how these influence learning outcomes. But what is still missing, is an overview and guideline for atomic design decisions that need to be taken into account when creating such a system. Based on a review of articles on pedagogical conversational agents, this paper provides an extension of existing classifications of characteristics as to include more fine-grained design aspects.

1. Introduction

Conversational Agents like Apple's Siri, Google Assistant or Amazon's Alexa are becoming more and more prevalent in a lot of people's lives. Not only can such voice-based digital assistants, or, shorter, conversational agents, be used at home, but Gartner [25] actually predicts, that by 2021, 25 percent of digital workers will be using such assistants on a daily basis. It is, also, likely, that this technology will diffuse into education context. Many applications (skills) for Alexa have already been developed to support formal and informal learning [10]. However, it remains open how such an agent should be designed to enhance the chances of positive learning outcome. What are the characteristics that support learning? What can the designers do and what atomic decisions should they take to make learning more effective when designing a pedagogical conversational agent? So far, the literature has been missing an overview and a consistent guidance on which aspects need to be considered and what their potential impacts are. Based on a review of articles on pedagogical conversational agents, this paper provides an overview of relevant aspects identified to impact the learning process.

Education is a field in which conversational agents have very widespread usage possibilities. At a lot of Universities students will visit a lecture on a topic repeatedly every week and will then go on to study the material by themselves. During this individual learning process questions about topics that were not sufficiently discussed during the lecture might come up. In such a situation, a human tutor would most likely be able to help out. But the problem in today's education systems is that the ratio between human tutors and students is not very balanced, which means that not every student is able to get the individual support he or she might require [9]. This is where pedagogical conversational agents become more relevant: they would be able to make up for the lack of available human tutors and give students the additional help they need. The following scenario will give an impression about how conversational agents might be used in education:

Lucy is a university student in the year 2021. She is working on getting a bachelor's degree in biology and therefore visiting a lot of different lectures. At the moment she is sitting at home and working on an individual homework assignment with e-learning content, which requires some statistical knowledge.

She can simply not remember what the difference between a median and a mean is, so she asks out loud: "Alice, what is the difference between a mean and a median?". The synthetic voice of the pedagogical conversational agent called Alice that is embedded in the e-learning system answers: "The mean is the average of all numbers and the median refers to the middle of all numbers when they are listed in numerical order".

This does not quite clarify the matter for Lucy, so she asks again: "Alice, show me an example of what a median is". In response to this, Alice opens a new window in Lucy's browser and shows a sequence of numbers, sorted in numerical order. In addition to the numbers, Alice's embodied representation also pops up next to the numbers. Using gestures and voice, Alice's virtual representation starts explaining what a median is. Thanks

Accepted for presentation at HICSS 2020

to this further explanation Lucy now knows how to solve the task in front of her.

In the next lecture the same e-learning script is referenced by the lecturer. Students are to split up into groups of two and discuss the topic of photosynthesis with each other while in the presence of the conversational agent Alice. Lucy and her discussion partner James don't really know where to start the conversation, so Alice steps in and suggests: "Why don't you start by discussing the component of carbon dioxide?". This short input is enough to launch Lucy and James into their discussion on photosynthesis.

This scenario shows some possible implementations of conversational agents in higher education, but the one thing that all described variations of the conversational agent Alice have in common, is that they serve as a version of a tutor. So far, a tutor, co-learner or lecturer have been human and have certain characteristics that make them a good (or bad) educator, but thanks to recent advances in the fields of artificial intelligence and natural language processing it has become possible to replace, or at least supplement, human educators with pedagogical conversational agents. However, this brings up the question of whether a computer in the form of a conversational agent also possesses the relevant characteristics that a good human educator has. Hence, the paper asks the following research question: *What aspects are relevant in the design of pedagogical conversational agents?*

The aim of this paper is to review existing classifications of (pedagogical) conversational agents and to extend these classifications with elements that are specifically relevant in the context of digital tutors in higher education. Here higher education refers to the education one receives at universities or equivalent establishments. We focus on higher education, because university courses involve a lot of individual at-home studying. And this is exactly where conversational agents could be potentially useful already now: when students are struggling to solve an individual task.

2. Related Work

In the design of pedagogical conversational agents many aspects can be considered. For one, there is the technology on which the agent is based. Depending on how it is implemented, the agent's usefulness and learner acceptance will be impacted. However, there are also other characteristics that affect the user experience and, thus, must be taken into account. In order to gain an overview over the possibilities in pedagogical conversational agent design, several classifications have been constructed. This section reviews both the technological background and the existing classifications of conversational agents.

2.1. Technology

When it comes to conversational agents, there are several interfering definitions. First, the term *conversational agent* refers to "software that interacts with its users through natural language" [6:1]. Natural language can be both via voice and text – in the latter case the agent would be classified as a chatbot [9]. This paper is about *pedagogical conversational agents*. These are conversational agents that are used in the context of education [9]. They might communicate via voice or text and might use various technologies, but they are learning and learner-oriented. A *pedagogical agent*, in contrast, is a virtual representation of a person that is used to recite educational information [5]. The difference between pedagogical agents and pedagogical conversational agents is that a pedagogical agent will merely hold a monologue, while a pedagogical conversational agent is able to engage in a dialogue with the learner [5].

Even though the differences between chatbots and voice-based conversational agents seem essential, the technology behind the stage is similar. On a very basic level, the only part in which speech-based agents and text-based chatbots differ form each other is the fact that a speech-based conversational agent will first need to transcribe the voice input into written text or some abstract representation thereof (speech recognition) and in the end convert the computed answer into spoken output (speech synthesis). Once the speech input has been transcribed using speech recognition software, the written text can be handled in the same way as pure text-based input [17, 18].

When it comes to the technologies that are employed for the implementation of conversational agent intelligence, there are two main options that one can choose from. For one, there are simple rule-based systems and on the other hand there exist more sophisticated and advanced self-learning approaches [6]. A rule-based conversational agent will perform pattern and keyword matching in order to understand the user input., be this voice or text. Since the rules for such systems need to be entered manually by the developers, designing such an agent can be very time-consuming, especially if a lot of different scenarios should be covered and understood by the bot. Although the defined rules can be arbitrarily simple or complex, a rule-based agent will usually fail to answer complex queries [18]. A rule-based pedagogical conversational agent should therefore be used in scenarios where straightforward questions with corresponding straightforward answers are expected. In the scenario above this would be the situation in which Lucy asks Alice what a median is.

In contrast to a rule-based conversational agent, a self-learning agent is based on machine. A self-learning agent can be used in more complex scenarios, when complex queries and complex answers are expected. In this context a self-learning agent refers to an initially corpus-based data-driven dialogue system that is later further augmented using data that is learned through its interactions with humans [17]. In order to be able to train such a conversational agent, one requires access to large amounts of conversational data. Since the developers have control over what data goes into the training of a conversational agent, it will ultimately learn exactly what it is supposed to learn. The used data can be cleaned and filtered before it is fed into the system, so as to avoid any incorrect information being taught. It is an entirely different case when the agent learns directly from its interactions with humans after it has already been deployed [17]. The same rigorous content control that goes into cleaning the initial data set also must be applied to any data that is used to train the agent after it has gone live. If this is not done correctly, the conversational agent could end up saying wildly inappropriate things and insulting its users, as it was the case with Microsoft Tay, who in 2016 learned inappropriate behavior from just 16 hours of interaction with human users.

Overall, the reliability of a conversational agent is a key issue. A speech-based conversational agent that never understands what the user is saying will lead to frustration and ultimately the discontinued use of the application. Similar to this is the problem of producing accurate and natural responses. A rule-based system that is poorly trained with too little keywords or missing patterns will likely produce inaccurate output which destroy the illusion that one is conversing with another human.

2.2. Classifying Conversational Agents

Conversational agents can be immensely useful in myriad of scenarios, ranging from education, as described in this paper, over task innovation and automation in organizations [18], all the way to making the average person's life easier by automating small tasks like setting a timer through Apple's Siri. But at the same time a lot of effort and fine-tuning goes into developing a successful conversational agent.

In order to support the design and implementation of conversational agents, a lot of platforms have emerged over the past years. To help organizations and developers decide on which platform they should use, Diederich et al. [6] have developed a taxonomy of conversational agent platforms. In their paper they describe how they analyzed 51 conversational agent platforms (of all sorts, not limited to pedagogical conversational agents) and have developed a morphological box consisting of 11 dimensions. Each of the 11 dimensions contains two to four characteristics (see Table 1).

Table 1. Morphological box of conversational agent platforms (Diederich et al. [6])

Parameter	Value			
Communication Mode	Text-based	Speech-based	Both	
Context	General-purpose	Domain specific		
Language	Single language	Multi language		
Intelligence	Rule-based	Self-learning		
Implementation	Programming	Modeling	Supervised learn.	Hybrid
Hosting	On-premise	Cloud	Both	
Pricing model	Usage-based	User-based	Instance-based	Free
Reporting	Without reporting	With reporting		
Sentiment detection	Without sentiment	With sentiment		
Enterprise integration	None	API	Pre-build interface(s)	
Platform integration	Single-platform	Cross-platform		

While Diederich et al. [6] have classified conversational agent platforms in general, Hobert and Meyer von Wolff [9] focused their research on pedagogical conversational agents. In their study, they have identified five dimensions, for each of which the characteristic that was most common in the reviewed literature has been highlighted (see Table 2). Their dimensions address the contextual aspects of how the agent is to be used rather than singular, atomic design decisions. Nevertheless, they offer a good overview of the tendencies to inform the decision concerning the setting, in which pedagogical conversational agents are likely to play a rising role.

Table 2. Morphological box of pedagogical agents (Hobert and Meyer von Wolff [9])

Parameter	Value			
Type	Messenger-like conversational agent	Embodied conversational agent		
Platform	Mobile-first	Web-based	Other	
Learning setting	Formal learning settings (e.g. at a university while attending a seminar)	Non-formal learning settings (e.g. self-study)		
Learning form	Isolated learning	Collective learning	Situated learning	Collaborative learning
Content	Single-topic learning content	Multiple-topic learning content		

Messenger-like conversational agents, as in chatbots, specifically developed for mobile platforms, seem to dominate over embodied conversational agents intended for web-based or other platforms. Pedagogical conversational agents that must be used in a specific context, like a university seminar, have a formal learning setting, whereas a pedagogical conversational agent that can be used "independently of a specific location, time or learning environment" [9:8] have a non-formal learning setting. The learning form describes whether the learner is dependent on location or other users. Notably, self-study in an isolated context (as indicated in the introduction) was identified as a promising direction of development [9]. Agents with single-topic learning content only support a specific learning scenario.

Multiple-topic learning content, in contrast, can be achieved by enabling lecturers to edit or add learning content themselves via control panels.

These two studies represent a good overview over the current state of, and potential of (pedagogical) conversational agents, but what is still missing, is a comprehensive view over the more detailed aspects of such agents. When designing a conversational agent for educational purposes there is not only the question about what platform should be used for development, or what type of context it should be intended for. In order to maximize the knowledge transfer, one needs to make very detailed decisions about what the agent should look like, and how it should interact with the learner.

Knowing that a conversational agent *can* be used as a tutor, does not mean that it *should* be used as such or that it can effectively replace human tutors. There are several characteristics that make humans good tutors. Edwards et al. [8] highlight three communication variables that have been proven to be important in instructional communication research: immediacy, credibility and teacher clarity. *Immediacy* refers to verbal and nonverbal gestures or cues that convey psychological closeness. Examples for such cues are smiling and nodding, or even using inclusive pronouns [8]. Pedagogical conversational agents can be either voice- or text-only or be present in an embodied form additionally. Since most of these immediacy cues require facial gestures, this trait can only satisfyingly be fulfilled by embodied conversational agents. But even if an embodied agent employs such immediacy cues, it would have to be ensured that these are used at appropriate timings. An agent that is constantly nodding or smiling, even if the conversational context does not warrant it, will not be taken seriously by the student and can therefore negatively impact the credibility of the conversational agent. The concept of *credibility* has already been referenced in the section on technology - if a conversational agent is poorly programmed, meaning that it does not understand the student correctly or consequently delivers the wrong answers, credibility and trust towards the system will suffer. The third relevant characteristic is *teacher clarity* in communicating information. There are a lot of instructor behaviors that can influence clarity, but the one in which conversational agents actually have an advantage over human tutors are vocalic cues [8]. The voice, pitch and accent of a machine agent can easily be altered to fit the ideal voice for clearly communicating information.

3. Method

To identify the relevant aspects of pedagogical conversational agents, according to the research question, we conducted a systematic literature study following vom Brocke et al. [1]. We defined the *scope* of the review to include scientific articles which describe the application of conversational agents in education and provide access to research outcomes addressing particular aspects of those agents. In doing so, we intend to integrate the knowledge and make it accessible to scholars and practitioners in a neutral manner. The *focal concept* addressed in the current study is a conversational agent designed for use in pedagogical manner, in the context of education. We particularly attend to such aspects and characteristics of those agents, which are subject to the design process and influence the learning outcome.

We conducted an exhaustive and selective *literature search*. We searched in the Google Scholar using the keyword "pedagogical conversational agent", which we identified earlier as the most adequate term (by informal consideration of background literature and by comparison to alternative queries like "education + conversational agent" or "education + digital assistant"). We used the default option from Google Scholar ("ALL"), which searches for all terms in the phrase and sorts them by their relevance (which gives preference to those which include the exact phrase, and, then, yields related results which may not include the exact phrase). We conducted two queries on Google Scholar: first, using the chosen keyword without any time range limitation, yielding 31'400 items without patents and citations; second, using the same keyword limited to papers published in 2015 or later, yielding 13'300 items without patents and citations. This additional search on the newest articles was conducted to include the most recent findings, which might have been otherwise remained uncovered because of their lower citation index. In late 2014, Microsoft released Cortana and in 2015 Amazon Echo (including the API for creation of Alexa Skills) was made available to the public, and in 2016 Apple released an API for Siri. We expected that those developments might have impacted the research in the chosen area. We took 50 top-most items from each query (sorted by relevance) to be considered for further processing. We employed a two-tier *evaluation* procedure to those 100 articles. First, the first author removed duplicates and determined the overall relevance of each article based on the title and abstract, which resulted in a set of 26 articles. Most of the articles were removed because they were not set in the context of higher education or were not focused on one or several specific parameters of the agent. Second, we reviewed the research context, independent and dependent variables, as well as the key findings, which reduced the number of relevant articles to 11. Through backward and forward search applied to those 11 articles, 9 more relevant articles were identified leading to the final set of overall 20 articles considered in the current study. All of them present and

evaluate the application of a pedagogical conversational agent and identify outcome-related aspects thereof. Based on these aspects we created a morphological box for pedagogical conversational agents (cf. Table 3).

The use case we approach in the current paper is not the typical classroom situation but rather an individual learning scenario, where a student or group of students study a topic on their own. For one, individual study in higher education context[18] is the trending use case according to Hobert and Meyer von Wolff [9] (cf. Learning Setting and Learning Form in Table 2). Furthermore, the conversational agents available in the consumer market (Alexa, Siri, etc.) are designed to support informal interaction with small groups of users rather than broadcasting to a large group: they react to specific questions, provide punctual answers, and facilitate interactive usage. All in all, the focus on informal and individual/small group education seems a timely and more urgent issue compared to visionary scenarios.

We use the insights from our literature study in addition to previous literature reviews to establish a morphological box. In particular, we reviewed the aspects of design varied, controlled for or identified as relevant for learning success in the selected literature. We then grouped the parameters into categories while using terms inspired by the papers. Morphological boxes have been used to study and suggest the design of socio-technical systems in IS and beyond [14, 15]. They help investigating the relationships in multi-dimensional, non-quantifiable problem complexes. Design of socio-technical systems where a computer takes on a social role belongs to this category of challenges. One can think of a morphological box as a multi-dimensional spatial cube, of parameters as dimensions of that cube, and of values as distinct positions on the dimension axes. Creating a morphological box makes the dependencies between the various aspects easier to grasp and process.

4. Results

The term "pedagogical conversational agents" dates back to the early 2000s. In that time, Nishida offered the vision of EgoChat, a virtualized ego designed to support group knowledge creation [13] and designated it as a "pedagogical conversational agents". The term was used scarcely over the subsequent 10 years involving only three distinct research groups. The early research yielded examples of chat-based games to support learning in specified, limited domains such as low-level math [19]. However, since 2010, the number of publications multiplied and the term settled.

Based on the found research articles the morphological box of pedagogical conversational agent characteristics seen in Table 3 was created. These characteristics are to be viewed as an addition to the characteristics defined in Table 1. All in all, eight parameters with two to four different values were identified in the considered literature. *Role*, *Function* and *Interaction Configuration* are the top-tier aspects, which we refer together as *Purpose Characteristics*. *Formality* and *Type of voice* describe the *Speech Characteristics*. And *Gender*, *Immediacy*, and *Gesturing* are the identified *Physical Characteristics*.

Table 3. Morphological box of pedagogical conversational agents in addition to Table 1.

	Parameter	Value			
Purpose	**Role** [11, 23]	Tutor		Co-learner	
	Function [20, 22, 24]	Source of information	Discussion help	Reflection tool	Task guidance
	Interaction Configuration [11, 21, 22]	Dialogue		Trialogue	
Speech	**Formality** [17]	Based on written dialogue corpora		Based on spoken dialogue corpora	
	Type of voice [2, 3]	Classic text-to-speech engine	Modern text-to-speech engine	Human voice	
Physical	**Gender** [12, 15]	Female	Male	Gender-neutral	
	Immediacy [8, 12, 15]	Reactive to user	Non-reactive to user	None	
	Gesturing [5, 8]	Deictic	Iconic	Beat	Metaphoric

4.1. Purpose Characteristics

Role: As mentioned in the introduction, pedagogical conversational agents can inhabit more than one role. For one, there is the *tutor*. A conversational agent tutor will help the student (or students) with self-study material inside and outside of the classroom. It can answer specific questions, help the student solve a difficult task, revise learned material and act as a moderator in an academically productive discussion. Then there is the conversational agent that acts in the role of a *co-learner*. A co-learner conversational agent will mimic a student peer and can either be programmed to be high-performing or low-performing and socially supportive or competitive [11]. The reason why one would want to learn with the help of an agent co-learner is to satisfy the learners' sociocultural needs [23]. The agent then acts as an activity partner and therefore provides opportunities for social interaction. Of course, there are further roles one could imagine. For instance, the third possible role is the pedagogical conversational agent that holds entire lectures in front of a classroom full of people. In this role the agent broadcasts the information within a large group. The focus of this paper, though, is on the interaction of pedagogical conversational agents with a single individual or a small group of students. Therefore, the scenario of a lecturer is not further discussed. The considered literature has so far identified the roles

of a tutor and a co-learner for the intended use cases. Those two roles complement each other.

Interaction Configuration: While the most conventional and basic form of interaction consists of one human interacting with one conversational agent, there are also several other ways in which pedagogical conversational agents can be included in learning. Apart from the traditional *dialogues* there also exist *trialogues* in which three entities are involved. The possible interaction configurations for these scenarios would be (1) one conversational agent representing the tutor and two human learners [21, 22], or (2) one conversational agent representing the tutor, another conversational agent in the role of a co-learner and one human learner [11].

The goal of such scenarios with three or more entities is to facilitate academically productive talk in small group discussions by delivering unsolicited interventions that are based on the academically productive talk framework. Regarding the first configuration, Tegos and Demetriadis [21] have found that by using such interventions, both individual and group learning outcomes could be improved. They encourage students to build on their prior knowledge and link this to new domains discussed in the course. Tegos et al. [22] built upon this and found that the impact on learning outcomes of the individual learner is even greater when the conversational agent does not address both learners simultaneously but uses directed intervention to target a specific learner. With regards to the second scenario Ju et al. [11] have shown that not only can a conversational agent influence learning outcome through acting in the role of a teacher or tutor, but also by representing a co-learner, more specifically a high-performing co-learner.

Function: Throughout the literature review, many different implementations of pedagogical conversational agents have been found. All of the them can be categorized into four different functions: source of information – F1, discussion helper – F2, reflection tool – F3, and guide through tasks – F4.

(F1) Source of information: This function describes the simplest form of pedagogical conversational agents. The learner can either ask for information explicitly or is presented with a specific piece of information based on the context he or she is in. An example for the latter would be a conversational agent that is situated in the context of a museum. As soon as the learner gets close to an exhibition piece, the conversational agent will provide the user with context specific information, without needing the user to ask a specific question.

*(F2) Discussion helper***:** When acting in the function of a discussion helper the pedagogical agent can take on both the role of a tutor, or then the role of a co-learner. A discussion helper in the context of small groups helps one or more human learners in their discussion on a specific topic by actively mentioning certain keywords to be considered and therefore guiding a discussion into the right and academically useful direction [22].

(F3) Reflection tool: A pedagogical conversational agent in the function of a reflection tool asks the learner content-related questions about what he or she has recently learned [20]. Such conversational agents can be used for revision of content that was introduced in a lecture, for example.

(F4) Guide through tasks: This type of tutor can be implemented for both simple problems and complex problems, although it will be most useful in complex problem-solving tasks. Winkler et al. [24] suggest that the agent should guide the user through the necessary steps to solve a specific (complex) problem. This kind of tutor is based on the constructivist learning theory, which states that a person will have the best learning outcomes if the learning process is interactive [24].

4.3. Speech Characteristics

Formality: As described in Section 2, the selection of the correct data set to train a conversational agent with is of utmost importance. Using data that is not cleaned, filtered or otherwise inspected before it is fed into the system can yield an agent that is biased and unintentionally rude to its users. But not only the behavior of the agent is affected by the choice of data. Also, the choice of words and level of formality in which the agent speaks are influenced, and this in turn has an impact on the students who are interacting with it. Serban et al. [17] explain that most available data sets are in the form of informal written dialogues between humans, the emphasis here being on the term *written*. These dialogues usually come in the form of movie scripts, forum posts, and micro-blogging platforms like Twitter, meaning that they were not transcribed from natural spoken dialogues, but intentionally written with the purpose of people reading the conversation, rather than speaking it. Looking at these data sets from a linguistic point of view, there are some concerns regarding the training of speech based conversational agents with them. Spoken conversations are usually less formal than written dialogues, have a different turn-taking structure, are highly interactive, multi-modal and socially situated. This means that in order to create a speech-based conversational agent that speaks in a more natural way, it is crucial to use actual spoken, multi-modal dialogue corpora.

Type of voice: But not only the choice of words and turn-taking style that a conversational agent is based on has an impact on learning. Also, the type of voice and how this voice is generated influences how effectively information can be delivered though speech-based conversational agents. Craig and Schroeder [3] have studied how a classic text-to-speech engine, modern text-to-speech engine and the human voice compare in a non-

interactive multimedia environment. The material to be learned by the participants consisted of visual images about the formation of lightning and 19 statements that were narrated by either a classical text-to-speech engine, a modern text-to-speech engine, or a recorded human voice. The classical speech-to-text engine is described as "While understandable to the listener, this voice had a digital quality with clipped or choppy production and no inflection" [3:5]. The modern text-to-speech engine "while still computer-generated without inflection or prosody, does not have the synthesized tone and has a smoother voice presentation" [3:5]. Interestingly, even though the modern text-to-speech engine uses no inflection or prosody, this study shows that there is no statistical difference between the learning outcomes, credibility, or cognitive efficiency measures when comparing the modern text-to-speech engines to the recorded human voice. When it comes to the perceived human-likeness and engagement, the actual human voice was rated significantly higher than the computerized voices.

Another interesting phenomenon was found by Craig and Schroeder [2]. In their study they paired an embodied virtual human with the same three variations of voice types and learning materials as used by Craig and Schroeder [3] (the voice types being classic text-to-speech engine, modern text-to-speech engine and recorded human voice). They showed that the scores in the learning transfer measures were higher when the virtual human was paired with the voice produced by the modern text-to-speech engine than when it used a human voice. Craig and Schroeder [3] in contrast, did not find this distinction when they investigated the voice effect without the presence of a virtual human. This demonstrates evidence on how the presence of virtual embodied human in learning contexts can influence other aspects of social interaction, including learning effect.

Even though these Craig and Schroeder's findings [2, 3] were collected in the context of virtual humans narrating the material to be learned in a non-interactive environment instead of using responsive conversational agents, the results might still have important implications for pedagogical conversational agents. In the case of speech-based pedagogical conversational agents it is not possible to record every possible response the agent could give using an actual human voice. Especially with self-learning agents becoming more prevalent, and therefore the answers the agent gives becoming more unpredictable, it is necessary to generate the speech output based on each individual response.

4.4. Physical Characteristics

The fact that the mere presence of an embodied virtual human can influence study results (as shown in the case of Craig and Schroeder [2] and Craig and Schroeder [3]) implies that it might be worth investigating the role of different physical characteristics of such embodied agents as well. Here it is important to note that the studies conducted in this field are all concerning pedagogical agents and not specifically pedagogical conversational agents. We have still chosen to include these characteristics in this paper because, based on the results of the following studies, we believe that the combination of advanced speech synthesis technology and 3D animation of virtual humans can lead to better learning results.

Gender: The most obvious feature of an embodied agent is its gender. It affects both its appearance as well as its voice. There seems to be conflicting evidence on whether or not gender plays a role in pedagogical (conversational) agents. In one study, the gender effect was studied using a pedagogical agent which simply narrates the learning material in the form of a video [16]. Here no effect of agent gender on learning could be established. Schroeder and Adesope [16] speculate that that the fact that the continuous flow of the video in which the agent was displayed did not allow the learners to apply unconscious thought processes such as gender stereotypes. The learners seem to have required the full extent of their working memory in order to keep up with the information presented in the video. The second explanation given, which states the lack of engagement with the agent to be the reason why gender did not affect learning outcomes, seems more plausible in light of the results by Krämer et al. [12]. They studied the situation, where the learner actually interacted with the agent. The study shows that the learner's performance was higher, when they interacted with an agent of the opposite gender. Overall, it seems that the effect of gender is moderated by the type of interaction or the immediacy.

Immediacy: The most notable difference between these two studies is that in the latter case the agent engages in a conversation with the learner, instead of just narrating the learning material, and builds rapport by displaying human-like behavior like smiling, nodding and blinking. The rapport building behavior of the agent (smiling, nodding, blinking) was automated based on the human users' audiovisual signals like voice and upper-body movements. The speech produced by the agent was manually controlled by the researchers during the experiment. Based on these two studies it looks like agent gender only makes a difference when the learner is actively engaged with the agent and not just listening to it over a video. The finding that agent gender makes a difference in learning outcomes would mean that a system containing the pedagogical conversational agent would need contain functionality to dynamically adjust traits of the agent to the learner. So, if the learner if female, the agent voice and embodiment should ideally be male. The importance of studying rapport building behavior in human-machine interaction is further stressed

by Edwards et al. [8]. Here the term immediacy or psychological closeness is used to describe the effect of verbal and nonverbal cues used by instructors in educational settings like classrooms or at-home tutoring.

Gesturing: The final physical characteristic is gesturing. When you think about having a conversation with another human, gesturing is a natural part of that conversation. We gesture in order to express meaning and importance, to show the shape of things and to point at objects and places. It also helps raise the clarity of communication, as mentioned by Edwards et al. [8]. In the case of pedagogical agents, the most prevalent types of gestures are deictic gestures. "Deictic gestures are performed to direct the spatial awareness of an individual" [5:194], or in other words, these are gestures that are used to point to some information. In their meta-analysis on pedagogical agent gesturing in multimedia learning environments Davis [5] found, that gesturing in pedagogical agents does indeed have an effect on learning outcomes (measured through near transfer and retention), though this effect is small. As a reason for this rather small effect they suggest that gesturing in pedagogical agents does not accurately represent the variety of gesturing types that are available to us in human-human conversations. Apart from deictic gestures we also use iconic, beat, and metaphoric gestures. This suggests that learning outcomes might be able to benefit from pedagogical agents that are programmed to use a more diversified arsenal of gestures, thereby mimicking human-human interaction more accurately.

5. Discussion

In this paper we have shown which characteristics are important when it comes to pedagogical conversational agents. A morphological box of pedagogical conversational agent characteristics was created and will now be discussed with respect to the related work. Additionally, directions for future research are discussed.

The existing morphological box of conversational agent platforms by Diederich et al. [6] aims to provide an overview of state-of-the-art platforms that support the development of conversational agents in general. While the aspects that they discussed might suffice when considering the development of a conversational agent, more dimensions need to be taken into account when the agent is to be used in the context of education. The paper by Hobert and Meyer von Wolff [9] describes what aspects surrounding pedagogical conversational agents, specifically, are to be considered in their design.

But according to the reviewed literature, pedagogical conversational agents require special attention to every single detail. Changing any aspect of such an agent can result in better, or potentially even worse, learning outcomes. The main contribution of this paper is an extension of the existing morphological box by Hobert and Meyer von Wolff [9] as to include more fine-grained design aspects of pedagogical conversational agents (see Table 4). Whereas the aspects of immediacy, credibility and teacher clarity [8] seem natural for humans, conversational agents can establish those only if their design considers a whole set of low-level features. For instance, whether the conversational agent appears credible to a learner depends not only on the content it provides but may also be affected by speech parameters. This requires the designers to consider the relevant aspects in order to overcome the limitation of the technology.

Table 4. Morphological box of pedagogical conversational agents

	Parameter	Value			
Technical	**Type**	Messenger-like conversational agent		Embodied conversational agent	
	Platform	Mobile-first	Web-based	Other	
Didactical	**Learning setting**	Formal learning settings (e.g. at a university while attending a seminar)		Non-formal learning settings (e.g. self-study)	
	Learning form	Isolated learning	Collective learning	Situated learning	Collaborative learning
	Content	Single-topic learning content		Multiple-topic learning content	
Purpose	**Role** [11, 23]	Tutor		Co-learner	
	Function [20, 22, 24]	Source of information	Discussion help	Reflection tool	Task guidance
	Interaction Configuration [11, 21, 22]	Dialogue		Trialogue	
Speech	**Formality** [17]	Based on written dialogue corpora		Based on spoken dialogue corpora	
	Type of voice [2, 3]	Classic text-to-speech engine	Modern text-to-speech engine	Human voice	
Physical	**Gender** [12, 15]	Female	Male	Gender-neutral	
	Immediacy [8, 12, 15]	Reactive to user	Non-reactive to user	None	
	Gesturing [5, 8]	Deictic	Iconic	Beat	Metaphoric

For one, the collected characteristics provide a basis for further research. In the reviewed papers most characteristics were studied in an isolated manner. By combining them, further insights into learning processes and outcomes supported by pedagogical conversational agents might be gained. On the other hand, the description of such detailed aspects can provide an additional guideline for the practical implementation of pedagogical conversational agents. Lecturers wishing to incorporate state-of-the-art technology into their lectures can base their design decisions on the overview provided in this paper. Of course, further layers could be added to the morphological box, for instance, to differentiate between the pedagogical, the strictly technological, and the operational aspects.

Four functions of conversational agents were shown, but not all of the functions have the same requirements when it comes to agent characteristics. Although Hobert and Meyer von Wolff [9] have shown that messenger-like agents are a lot more widespread than embodied conversational agents, this paper argues that, depending on the situation, an embodied agent could be a lot more effective than mere text-based agents. Some characteristics, like immediacy, can be conveyed most effectively through embodied conversational agents. For each of the functions F1-F4 this paper subsequently discusses whether a text-based agent would suffice, or whether a speech-based, possibly embodied agent is needed for maximum effectiveness.

According to the media richness theory, the richness of a medium used to communicate information should be proportional to the complexity of the communication task [4]. When it comes to conversational agents in the function of a simple source of information (F1), we believe that no embodiment is needed. The content presented by such an agent is usually fairly simple and needs no physical characteristics like gesturing in order to deliver information effectively. Whether or not speech is needed depends on the specific use case. For one, when the interaction with the agent happens spontaneously while the learner is, for example, working on a homework task at home, then a speech-based interface would probably be more useful. This way the interaction could be more spontaneous, and the learner would not have to go through the physical effort of typing his or her question into the computer or mobile device. Should the interaction happen in a public place, on the other hand, the conversation should be text-based, because studies show that people don't feel comfortable talking to their mobile devices in public [7].

Pedagogical conversational agents in the functions of "discussion helper" (F2), "reflection tool" (F3) and "guidance through tasks" (F4) all require a physically represented conversational agent. In the case of the discussion helper this is so, because the agent is used in the presence of at least one additional entity, so all in all at least three entities are involved in the discussion. It does not even matter whether the third party is human or a second conversational agent. When a person is involved in a discussion, he or she needs a reference point to look at while listening or speaking to another entity. A similar principle applies to the conversational agent as a reflection tool or a guide for complex problem-solving tasks. Since the dialogue in these scenarios is usually longer than just a few separate utterances, we assume that the learner needs something to look at while he or she is talking. Here immediacy traits come into play [8].

Hobert and Meyer von Wolff [9] state in their paper that using conversational agents in the context of education is not a big change for most students, seeing as they are already used to speaking to machines like Apple's Siri or Amazon's Alexa. Also, messenger-like text-based systems, so-called chatbots, are frequently used by students in their everyday lives. But whether experience with smart personal assistants and chatbots will raise the acceptance of pedagogical conversational agents is debatable. The nature of the tasks that would be fulfilled using pedagogical conversational agents is far more complex than what can be done using traditional (non-educational) conversational agents and chatbots. Further studies are needed in order to find out how easily students are willing to welcome a pedagogical conversational agent into their study habits.

6. Conclusion

In this paper, we discuss how conversational agents can be designed and which design aspects influence the learning outcome. The results do not come without limitations. First of all, the considered literature was selected based on a limited set of search queries and a narrow focus of interest. Extending the queries to include alternative terms like education, learning, studying, as well as robot, avatar, etc. could identify more relevant aspects to be considered. Also, he focus on an individual learning scenario that complements a classroom interaction limits the applicability – considering learning theories and identifying sets of similar learning scenarios could help determine the applicability of the results, and select further areas of inquiry. Second, we used Google Scholar to search for literature, which makes the results only partially reproducible, because of the proprietary search models and unclear relevance sorting. On the other hand, Google Scholar was chosen, because of its built-in features like automatic use of synonyms, which make the exploration easier. Third, the external validity of the review is compromised by the publication bias.

The current study systematizes insights concerning the design and impact of pedagogical conversational agents. The resulting morphological box supports researchers studying use of conversational agents in education at dividing and focusing their studies. It, also, raises open questions on how the identified characteristics influence each other and what is their relation to each other. The analyzed literature suggests some dependencies (e.g., gender and immediacy), but they require a more extensive and systematic approach. The resulting overview can, also, be used as a guideline for developers of pedagogical conversational agents. They obtain an overview of what aspects require special attention beyond the technical aspects. Overall, the conducted review and the resulting insights have practical implications and offer potential for further research.

7. References

[1] vom Brocke J, Simons A, Niehaves B, Riemer K (2009) Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: Proc. Europ. Conf. Information Systems, AIS.

[2] Craig SD, Schroeder NL (2017) Reconsidering the voice effect when learning from a virtual human. Computers & Education 114:193–205.

[3] Craig SD, Schroeder NL (2018) Text-to-Speech Software and Learning: Investigating the Relevancy of the Voice Effect. Journal of Educational Computing Research 0:1-15.

[4] Daft RL, Lengel RH (1986) Organizational Information Requirements, Media Richness and Structural Design. Management Science 32:554–571

[5] Davis RO (2018) The impact of pedagogical agent gesturing in multimedia learning environments: A meta-analysis. Educational Research Review 24:193–209.

[6] Diederich S, Brendel AB, Kolbe LM (2019) Towards a Taxonomy of Platforms for Conversational Agent Design. In: Proc. Intl. Conf. Wirtschaftsinformatik, Univ. of Siegen.

[7] Easwara Moorthy A, Vu K-PL (2015) Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. Intl. J. of Human-Computer Interaction 31:307–335.

[8] Edwards C, Edwards A, Spence PR, Lin X (2018) I, teacher: using artificial intelligence (AI) and social robots in communication and instruction. Communication Education 67:473–480.

[9] Hobert S, Meyer von Wolff R (2019) Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents. In: Proc. Intl. Conf. Wirtschaftsinformatik, Univ. of Siegen.

[10] Incerti F (2017) Amazon Echo: Emerging technology for formal or informal learning? In: Proc. Association for the Advancement of Computing in Education (AACE) Intl. Conf., pp 1627–1633

[11] Ju W, Nickell S, Eng K, Nass C (2005) Influence of Colearner Agent Behavior on Learner Performance and Attitudes. In: Extended Abstracts on Human Factors in Computing Systems, ACM.

[12] Krämer NC, Karacora B, Lucas G, Dehghani M, Rüther G, Gratch J (2016) Closing the gender gap in STEM with friendly male instructors? On the effects of rapport behavior and gender of a virtual agent in an instructional interaction. Computers & Education 99:1–13.

[13] Nishida T (2002) A traveling conversation model for dynamic knowledge interaction. Journal of Knowledge Management.

[14] Ritchey T (2002) Modelling Complex Socio-Technical Systems Using Morphological Analysis. Adapted from an address to the Swedish Parliamentary IT Commission, Stockholm, http://www.swemorph.com/pdf/futures.pdf

[15] Ritchey T (1998) General Morphological Analysis. A general method for non-quantified modelling. In: Proc. EURO Conf. Operational Analysis, Swedish Morphological Society.

[16] Schroeder NL, Adesope OO (2015) Impacts of Pedagogical Agent Gender in an Accessible Learning Environment. J. Educational Technology & Society 18(4):401-411.

[17] Serban IV, Lowe R, Henderson P, Charlin L, Pineau J A Survey of Available Corpora for Building Data-Driven Dialogue Systems: The Journal Version. Dialogue & Discourse 9(1):1-49.

[18] Shridhar K (2017) Rule based bots vs AI bots. In: #WeCoCreate. https://medium.com/botsupply/rule-based-bots-vs-ai-bots-b60cdb786ffa. Accessed 18 May 2019

[19] Sjödén B, Silvervarg A, Haake M, Gulz A (2011) Extending an Educational Math Game with a Pedagogical Conversational Agent: Facing Design Challenges. In: De Wannemacker S, Clarebout G, De Causmaecker P (eds) Interdisciplinary Approaches to Adaptive Learning. A Look at the Neighbours. Springer Berlin Heidelberg, pp 116–130

[20] Song D, Oh EY, Rice M (2017) Interacting with a conversational agent system for educational purposes in online courses. In: Proc. Intl. Conf. on Human System Interactions (HSI). IEEE.

[21] Tegos S, Demetriadis S (2017) Conversational Agents Improve Peer Learning through Building on Prior Knowledge. J. Educational Technology & Society 20(1):99-111.

[22] Tegos S, Demetriadis S, Papadopoulos PM, Weinberger A (2016) Conversational agents for academically productive talk: a comparison of directed and undirected agent interventions. Intl. J. of Computer-Supported Collaborative Learning 11:417–440.

[23] Veletsianos G, Russell GS (2014) Pedagogical Agents. In: Spector JM, Merrill MD, Elen J, Bishop MJ (eds) Handbook of Research on Educational Communications and Technology. Springer, New York, NY, pp 759–769.

[24] Winkler R, Bittner E, Söllner M (2019) Alexa, Can You Help Me Solve That Problem – Understanding the Value of Smart Personal Assistants as Tutors for Complex Problem Tasks. In: Proc. Intl. Conf. Wirtschaftsinformatik, Univ. of Siegen.

[25] Gartner Predicts 25 Percent of Digital Workers Will Use Virtual Employee Assistants Daily by 2021. https://www.gartner.com/en/newsroom/press-releases/2019-01-09-gartner-predicts-25-percent-of-digital-workers-will-u. Accessed 20 Mar 2019.